EXHIBIT G

Filed: February 10, 2021

Filed on behalf of
Patent Owner Advanced Cluster Systems, Inc.
By: Jon W. Gurka
Ted M. Cannon
Cheryl T. Burgess
KNOBBE, MARTENS, OLSON & BEAR, LLP
2040 Main Street, Fourteenth Floor
Irvine, CA 92614
Tel.: (949) 760-0404
Fax: (949) 760-9502
Email: BoxZTANNL.017LP2@knobbe.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

NVIDIA CORPORATION,
Petitioners,

v.

ADVANCED CLUSTER SYSTEMS, INC.
Patent Owner

Case No. IPR2021-00019
U.S. Patent 10,333,768

PATENT OWNER PRELIMINARY RESPONSE

TABLE OF CONTENTS

Page No.

I. INTRODUCTION AND SUMMARY OF ARGUMENT 1

II. CLAIM CONSTRUCTION 6

A. "a mechanism for the nodes to communicate . . . with each other using a peer-to-peer architecture" (all challenged claims) 6

1. The "wherein" clause refers back to the same "mechanism for the nodes to communicate . . . with each other using a peer-to-peer architecture" introduced earlier in the claim. 7

2. "peer-to-peer architecture" 8

B. "cluster node module" (claim 4) 12

III. THE PETITION DOES NOT ESTABLISH A REASONABLE LIKELIHOOD THAT ANY CHALLENGED CLAIM IS UNPATENTABLE 17

A. The Petition does not establish that the Distributed Maple Code references are prior art printed publications. 17

B. The Petition does not establish claims 1, 4-10, 18-22, 24-25, 30-31, or 33-34 are unpatentable. 20

1. The Petition does not show that the prior art discloses the "peer-to-peer architecture" limitation. 20

2. The Petition does not show that the prior art discloses communication of "user instructions" using the "peer-to-peer architecture." 24

TABLE OF CONTENTS
(cont'd)

Page No.

3. The Petition does not show that the prior art discloses "wherein the third node comprises a third hardware processor with a plurality of processing cores, wherein the third node is configured to receive the result of the first mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node." 26

C. The Petition does not establish claims 4, 6-10, 30, or 31 are unpatentable. 29

D. The Petition improperly relies on alleged public use evidence. 31

E. Objective evidence confirms the non-obviousness of the challenged claims. 34

1. Patent Owner's SEM™ and SET™ products met a long-felt but previously unmet need. 34

2. SEM™ and SET™ achieved surprising results. 38

3. SEM™ and SET™ successfully addressed the need, where others had failed. 40

4. SEM™ and SET™ received industry praise. 40

5. SEM™ and SET™ were met by initial skepticism. 41

6. SEM™ and SET™ embody the challenged claims. 42

7. There is evidence that Petitioner copied the claimed invention. 47

should reject Petitioner's attempt to go beyond the "patents and printed publications" upon which an IPR petition must be based, and find that Petitioner failed to meet its initial burden to show a reasonable likelihood it would prevail with respect to at least one claim.

E. Objective evidence confirms the non-obviousness of the challenged claims.

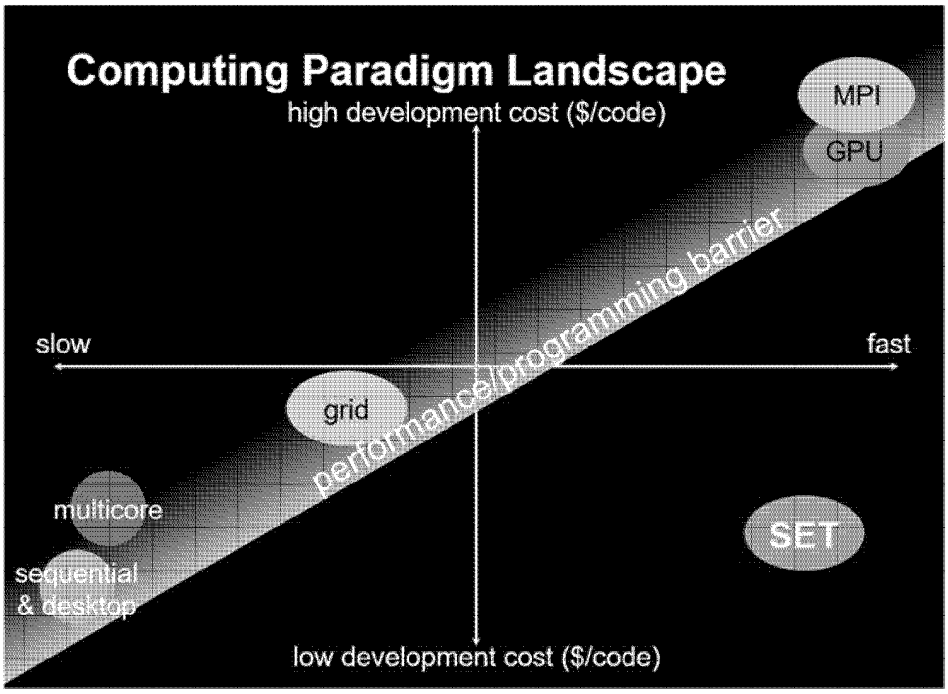
Objective evidence strongly supports the non-obviousness of the challenged claims, including meeting a long-felt, unmet need; the claimed invention's surprising results; failure of others to meet this need; industry praise; initial skepticism; and copying of the claimed invention. *See, generally*, Ex. 2017; Ex. 2018 at 27-28; Ex. 2019 at 2; Ex. 2020. When present, objective evidence of non-obviousness must be considered. *Apple Inc. v. Samsung Elecs. Co.*, 839 F.3d 1034, 1048 (Fed. Cir. 2016), *cert. denied*, 138 S. Ct. 420 (2017).

1. Patent Owner's SEM™ and SET™ products met a long-felt but previously unmet need.

Parallel computation on cluster computers increases the speed at which computations can be performed. Indeed, some particularly complex computations require parallel processing because they simply use too much memory or other computational resources for a single processor to effectively handle. Parallel programming using traditional parallel-computing architectures was notoriously difficult, time-consuming, and expensive. Ex. 2027 ¶¶ 13-18 (citing Exs. 2014,

2024-2025) (collecting quotes); Ex. 2007 ¶ 8; Ex. 2008 ¶ 21. To achieve the highest performance advantages of cluster computing requires asynchronous, peer-to-peer message passing capability between the nodes, such as by using a message-passing paradigm like MPI. Ex. 2027 ¶ 13. This MPI-style parallelized code is the highest performing, but also traditionally the most difficult to implement. *Id.*

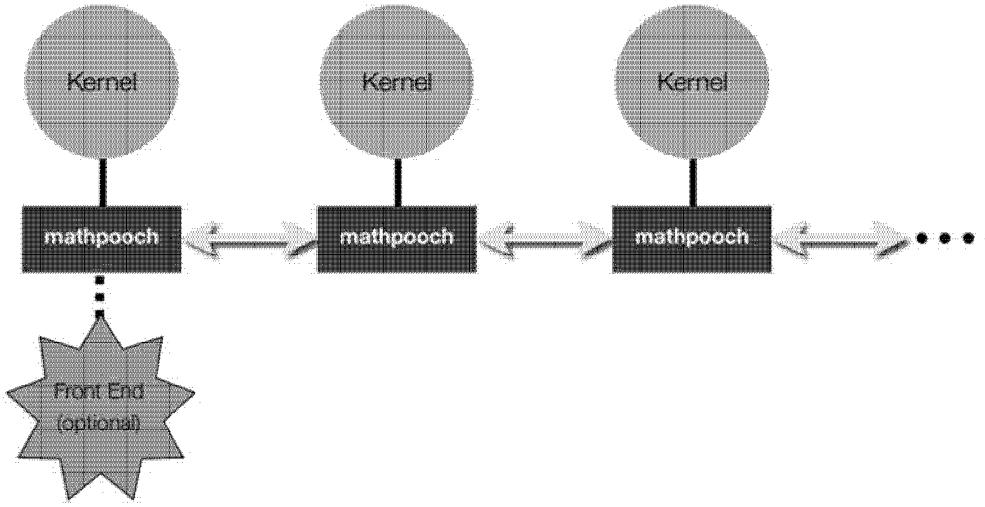
Parallel code was typically generated by converting serial code. Ex. 2008 ¶ 25. A programmer would take serial code and then "break" it into pieces for execution on each of the nodes in a parallel computer. *Id.* The programmer then added to each piece the message passing to accomplish inter-nodal communication. Using traditional parallel-computing architectures, only an experienced programmer with specialized parallel programming expertise could develop this type of parallelized code. Ex. 2027 ¶¶ 13-18; Ex. 2007 ¶ 8; Ex. 2008 ¶ 21. And while this process was difficult and time-consuming even for experienced parallel programmers—it was practically impossible for ordinary programmers. Ex. 2027 ¶¶ 13-18; s*ee, generally* Ex. 2017. The relationship between traditional parallel computing performance and the difficulty of the parallel programming is depicted below:



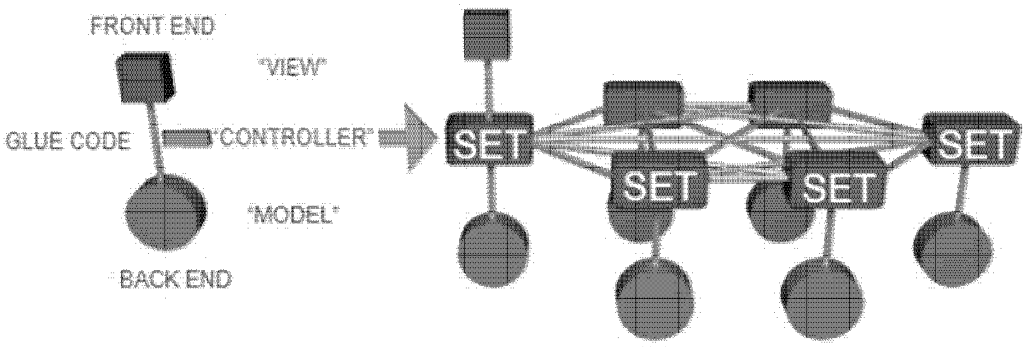
Ex. 2015 at 36. ACS dubbed the inability to achieve results comparable to the highest possible parallel computing performance with a low level of difficulty "the performance/programming barrier." *Id.*; *see also* Ex. 2027 ¶ 19. This posed an "insurmountable barrier for many organizations" that otherwise would have taken advantage of cluster computing. Ex. 2027 ¶ 17 (citing Ex. 2025). Thus, there was a long-felt but unmet need for a way to unlock the performance advantages of cluster computing without requiring specialized expertise or excessive time, effort, and cost (*i.e.*, a need to break the performance/programming barrier). Ex. 2027 ¶¶ 10-19; Ex. 2007 ¶¶ 8, 11; Ex. 2008 ¶¶ 16-23.

Patent Owner developed a cluster-computing architecture, used in its SEM™ and SET™'s products, that met this long-felt, unmet need. Ex. 2027 ¶¶ 20-30. The

new architecture interposed a communication layer between the front end user interface and the kernels running on each node of the cluster, or back end. Below is a diagram from one of the first SEM™ manuals that illustrates this architecture:



Ex. 2011 at 4. Below is a diagram from one SET™ datasheet that illustrates this architecture:



Ex. 2016 at 1.

SEM™ and SET™'s cluster-computing architecture enables parallel execution of Mathematica and more general applications, respectively, with a high

level of performance but without extensive specialized programming expertise, or the excessive investment of time and effort demanded by traditional parallel-computing architectures. Ex. 2027 ¶¶ 20-30; Ex. 2007 ¶¶ 10, 12 (discussing SEM™); Ex. 2008 ¶¶ 25-28 (discussing SET™). The unique SEM™ and SET™ cluster-computing architecture embodied by the challenged claims is the reason that SEM™ and SET™ were able to meet the long-felt but previously unmet need. Ex. 2027 ¶¶ 24, 27-29; *see also* Ex. 2001 ¶¶ 83-88; *infra* Section III(E)(6). This fact weighs heavily in favor of non-obviousness. *WBIP, LLC v. Kohler Co.*, 829 F.3d 1317, 1332 (Fed. Cir. 2016) ("Evidence of a long felt but unresolved need tends to show non-obviousness because it is reasonable to infer that the need would have not persisted had the solution been obvious.").

2. SEM™ and SET™ achieved surprising results.

SEM™ and SET™'s superior performance and ease of use were unexpected. Ex. 2027 ¶¶ 38-41. As discussed above, traditional parallel-computing architectures presented a performance/programming barrier. *See supra* Section III(E)(1). But the SEM™ and SET™ cluster-computing architecture enabled parallel performance approaching the highest speeds possible using traditional parallel-computing architectures in an unexpectedly easy, fast, and inexpensive manner.

As one example, SEM™ performed better than gridMathematica in terms of raw performance using a parallel computing benchmark. SEM™ and

gridMathematica exhibited comparable performance when both applications passed small messages. Ex. 2027 ¶ 39 (citing Ex. 2010). SEM™ substantially outperformed gridMathematica for larger messages, even when gridMathematica used more nodes. *Id.* (citing Ex. 2010). This performance advantage was unexpected, but even more surprising was that SEM did not require specialized expertise to obtain the advantage. Ex. 2027 ¶¶ 39, 41; Ex. 2007 ¶¶ 9-10, 12.

Similarly, the application of SET™'s architecture to large, complex applications facilitated parallelization of those applications in far less time and with less effort than would have been required using traditional parallel-computing architectures. For example, it might take a skilled parallel programmer 12 to 18 months to rewrite the serial code for a typical large mainstream application as parallel code using traditional parallel-computing architectures. Ex. 2027 ¶ 40 (citing Ex. 2016 at 3). But using SET™'s architecture, ACS was able to parallelize Wolfram Research's Mathematica in 1 engineer-month, Apple's HD QuickTime Exporter in 1 engineer-month, and Equalis's Scilab in 2.5 engineer-months. *Id.* (citing Ex. 2016 at 3).

That SEM™ and SET™ achieved such surprising results because of their unique architecture embodied by the claims weighs heavily in favor of non-obviousness. *Circuit Check Inc. v. QXQ Inc.*, 795 F.3d 1331, 1337 (Fed. Cir. 2015)

(finding that the fact a skilled artisan would have been surprised supported the jury's finding that objective evidence of non-obviousness existed.)

3. SEM™ and SET™ successfully addressed the need, where others had failed.

SEM™ and SET™ succeeded where others failed. Ex. 2027 ¶¶ 36-37; Ex. 2007 ¶¶ 29-31; *see also* Ex. 2007 ¶ 11. Before SEM™ and SET™, no one had succeeded at breaking the performance/programming barrier. *See supra* Section III(E)(1). Others had tried to do so by developing automatic parallelizers or universal compilers that would take serial object code as input and output object parallel code. Ex. 2027 ¶ 37; Ex. 2008 ¶ 30. However, none of these attempts achieved performance comparable to what could be achieved using traditional parallel-computing architectures. Ex. 2027 ¶¶ 37, 53 (citing Ex. 2021 at 2); Ex. 2008 ¶¶ 30. This strongly supports non-obviousness. *Uniroyal, Inc. v. Rudkin-Wiley Corp.*, 837 F.2d 1044, 1054 (Fed. Cir. 1988) ("[F]ailure of others to provide a feasible solution to a long standing problem is probative of nonobviousness.").

4. SEM™ and SET™ received industry praise.

Those familiar with SET™ and SEM™ praised them. Ex. 2027 ¶¶ 42-47. For example, PIMCO's Managing Director of Quantitative Modeling and Risk Analytics, Vineer Bhansali, Ph.D., stated that he "found SEM[TM] to be ***very efficient*** in terms of ***stability*** and use for financial engineers who do not have the time to

optimize load balancing issues but want to focus on modeling." Ex. 2007 ¶ 12; Ex. 2027 ¶¶ 42-43 (citing Ex. 2018 at 3) (emphasis added).

Former professor at the Tokyo Institute of Technology, Yuko Matsuda, stated "I CAN endorse SEM[TM] with pride. I would emphasize the ***flexibility*** and ***scalability*** of SEM[TM]. And SEM[TM] can make writing Mathematica programs even more ***flexibl***e than before." Ex. 2027 ¶ 44 (citing Ex. 2018 at 4) (emphasis added). Professor Matsuda went on to write a white paper discussing SEM™, in which he praised SEM™ stating, "[f]or Mathematica users, SEM[TM] stands in an advantageous position compared to the Parallel Computing Toolkit (does not use MPI and adopts a Wait/Queue protocol) by Wolfram." *Id.* ¶ 45 (citing Ex. 2023 at 2).

This praise supports non-obviousness. *WBIP*, 829 F.3d at 1334 (industry praise of "a product which embodies the patent claims weighs against an assertion that the same claim would have been obvious"); *see also Institut Pasteur & Universite Pierre Et Marie Curie v. Focarino*, 738 F.3d 1337, 1347 (Fed. Cir. 2013).

5. SEM™ and SET™ were met by initial skepticism.

Because so many had tried but failed to break the performance/programming barrier, many were skeptical that the SEM™ architecture could truly enable the high performance of cluster computing without requiring specialized expertise or excessive time and effort. Ex. 2007 ¶ 8; Ex. 2008 ¶ 32.

Experts were equally skeptical of SET™. For example, the Department of Energy ("DOE") considered three SET™-based grant applications submitted by Patent Owner. The DOE's industry expert reviewers expressed skepticism that SET™ would work. Ex. 2027 ¶¶ 48-54 (citing Exs. 2019, 2021-22). One reviewer stated "[t]he applicant . . . appears to offer an unverified solution to mak[e] parallel computing 'easy'. This reviewer . . . is ***highly skeptical of the results claimed in the proposal***. Without quantitative proof that SET[TM] provides the results, ***I cannot believe that the solution is viable***." Ex. 2027 ¶ 52 (Ex. 2019 at 2) (emphasis added); *see also id.* ¶ 53 (citing Ex. 2019 at 3-4) (misinterpreting SET™ to be an auto parallelizer).

This initial skepticism highlights the non-obviousness of the claimed invention. *WBIP*, 829 F.3d at 1335-36 ("If industry participants or skilled artisans are skeptical about whether or how a problem could be solved or the workability of the claimed solution, it favors non-obviousness."). The initial skepticism was entirely unwarranted given the technical success of SEM™ and SET™.

6. SEM™ and SET™ embody the challenged claims.

There is sufficient nexus between the objective evidence related to SEM™ and SET™ and the challenged claims to accord substantial weight to the objective evidence of non-obviousness. Patent Owner's SEM™ and SET™ products practice at least the challenged independent claims. *See generally* Exs. 2009-2016. Dr. Singh

and Dr. Dauger's testimony provide a more detailed analysis of how SEM™ and SET™ practice the challenged claims. Ex. 2001 ¶¶ 74-79; Ex. 2028; Ex. 2029; *see also* Ex. 2027 ¶¶ 31-35.

For example, with respect to claim 1, Dr. Singh explains that the contemporaneous documents show that a cluster running SEM™ or SET™ embodies the fundamental claimed architectural improvements over the prior art. The following figure depicts aspects of a SET™ cluster.

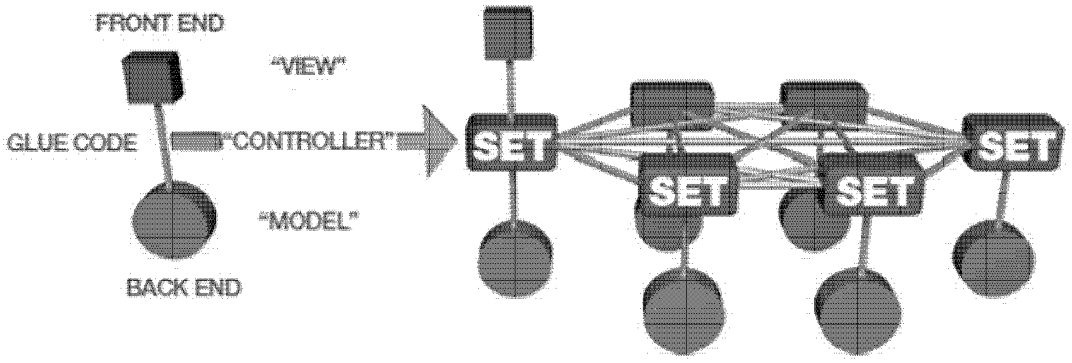


Figure 1. MVC (Left) and how SET is integrated into MVC (Right)

Ex. 2016 at 1. Dr. Singh explains that the figure and accompanying text of the documents show that the blue "SET" rectangles are "a mechanism for the nodes to communicate . . . using a peer-to-peer architecture," the gray "FRONT END" square is a "user interface," and the orange "BACK END" circles are "kernels." Ex. 2001 ¶ 77; Ex. 2028 at 3-5.

The documents also show that the blue "SET" rectangles satisfy the "peer-to-peer architecture" construction set forth above. Every "SET" block is connected to

every other "SET" block such that the nodes cooperate to establish intercommunication between the nodes and use MPI to "handle[] low-level message passing internally." Ex. 2013 at 2; Figure 1. Further, the "SET" blocks handle "MPI low-level tasks such as communication and synchronization between processes, data partitioning and distribution, mapping of processes onto processors, and input/output of data structures" and provide "high-level calls for common parallel tasks as Collective communication, Divide and conquer execution and data generation, guard-cell management, element management, and more." Ex. 2016 at 1. There is no indication in the SET documents that tasks and data are required to go through a central server or master node. Dr. Singh concluded that each of the "SET" blocks "can communicate tasks and data with other nodes without the tasks and data being required to go through a central server or master node," and, thus, includes the "peer-to-peer architecture" required by the claim. Ex. 2001 ¶¶ 77, 79; Ex. 2029 (in particular at 3-5); *see also* Ex. 2027 ¶¶ 26, 33-35. A similar analysis shows that SEM™ also embodies the challenged independent claim. Ex. 2001 ¶¶ 76, 79; Ex. 2027.

Dr. Singh also explains that the documents show that a cluster running SEM™ or SET™ embodies all claim limitations of the challenged independent claim, including:

- A computer cluster with a plurality of nodes (*e.g.*, nodes of SEM or SET cluster);
- each of the plurality of nodes comprises a hardware processor;
- one or more of the nodes are configured to receive a command to start a cluster initialization process for the computer cluster (*e.g.*, SETInitialize);
- each of the nodes is configured to access a non-transitory computer-readable medium (*e.g.*, memory) comprising program code for a single-node kernel that, when executed, is capable of causing the hardware processor to evaluate mathematical expressions (*e.g.*, SEM's Mathematica kernels or SET's parallelized application back end code);
- a mechanism for the nodes to communicate results of mathematical expression evaluation with each other using a peer-to-peer architecture (*e.g.*, SEM module or SET module);
- a first node comprising a first hardware processor configured to access a first memory comprising program code for a user interface (*e.g.*, a front end) and program code for a first single-node kernel, the first single-node kernel configured to interpret user instructions and distribute calls to at least one of a plurality of other nodes for execution (*e.g.*, SEM's Mathematica kernels or SET's parallelized application back end code);

- a second node comprising a second hardware processor with a plurality of processing cores (*e.g.,* a multicore processor), wherein the second node is configured to receive calls from the first node, execute at least a first mathematical expression evaluation, and communicate a result of the first mathematical expression evaluation to a third node (*e.g.,* when executing a parallel Fourier transform);
- a third node comprising a third hardware processor with a plurality of processing cores cores (*e.g.,* a multicore processor), wherein the third node is configured to receive the result of the first mathematical expression evaluation from the second node, execute at least a second mathematical expression evaluation using the received result, and communicate the result of the second mathematical expression evaluation to the first node (*e.g.,* when executing a parallel Fourier transform);
- the first node is configured to return the result of the second mathematical expression evaluation to the user interface (*e.g.,* when the parallel fourier transform is complete);
- one or more of the nodes are configured to: accept user instructions; after accepting user instructions (e.g., from the front end), communicate at least some of the user instructions using the mechanism for the nodes to communicate with each other (*e.g.,* SEM module or SET module); and after

communicating at least some of the user instructions using the mechanism, communicate at least some of the user instructions to one or more single-node kernels (*e.g.*, SEM's Mathematica kernels or SET's parallelized application back end code).

Ex. 2001 ¶¶ 74-79; Ex. 2028; Ex. 2029; *see also* Ex. 2027 ¶¶ 31-35.

Each of the objective indicia of non-obviousness results from the SEM™ and SET™ architecture embodied by the challenged claims. SEM™ and SET™'s claimed architecture allowed those products to meet the long-felt but previously unmet need by breaking the performance/programming barrier and enabling high-performance cluster computing without the specialized expertise or excessive time or effort required by traditional parallel-computing architectures. Ex. 2027 ¶¶ 21-31. Other parallel-computing architectures failed to meet the long-felt, unmet need precisely because they lacked SEM™ and SET™'s claimed architecture. *Id.* ¶ 37. The exceptional performance and other surprising results achieved by SEM™ and SET™ were also due to the claimed architecture. *Id.* ¶¶ 39-42. SEM™ and SET™'s claimed architecture is also directly responsible for the industry praise for the products and the products' ability to prove initial skepticism wrong. *Id.* ¶¶ 47-48.

7. There is evidence that Petitioner copied the claimed invention.

At a meeting conducted under a non-disclosure agreement in November 2012, Patent Owner explained to Petitioner the unique cluster-computing architecture of

the claimed invention embodied in the SET™ product. *Id.* ¶ 56. After the meeting, Patent Owner sent Petitioner, in confidence, an email attaching a specially tailored data sheet showing how Patent Owner's patented technology could complement Petitioner's general-purpose GPUs ("GPGPUs") and provide a communications infrastructure for direct all-to-all communications between each GPU. *Id.* ¶¶ 57-59. The datasheet included the following figure showing how SET™'s architecture could complement Petitioner's Tesla GPUs.

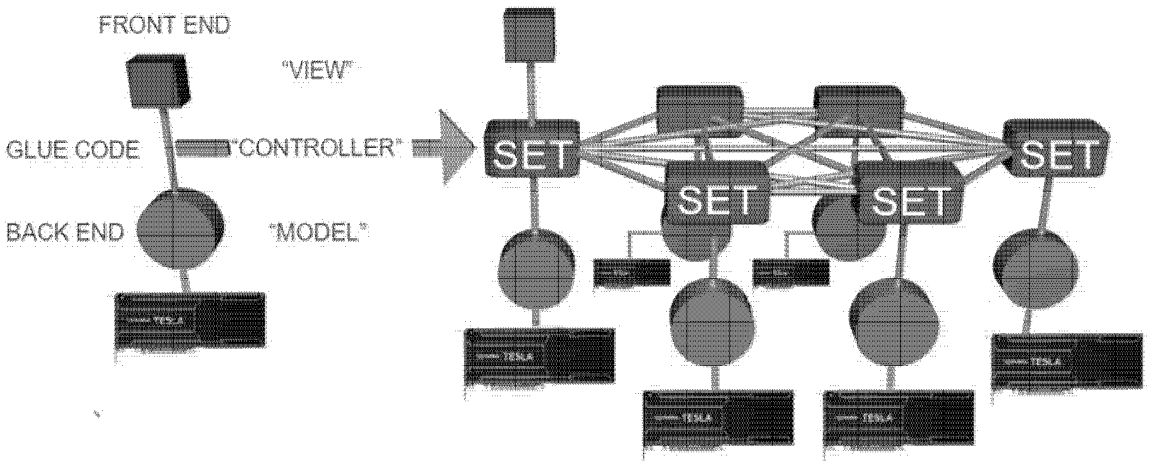


Figure 3. SET is a Perfect Companion to GPGPUs

Ex. 2020 at 4. The datasheet was marked with U.S. Patent Nos. 8,082,289 and 8,140,612. Ex. 2027 ¶ 58; Ex. 2020 at 4.

Petitioner then copied the claimed invention by incorporating the claimed architecture into Petitioner's GPGPUs in the manner described by the datasheet. Petitioner eventually named its GPU interconnect architecture, which uses the

claimed architecture, NVLink™. Petitioner then released a new version of CUDA[7] that, for the first time, added support for NVLink™. Ex. 2027 ¶ 59.

Accordingly, Petitioner's copying of the claimed invention weighs heavily in favor of non-obviousness. *Windsurfing Int'l, Inc. v. AMF, Inc.*, 782 F.2d 995, 1000 (Fed. Cir. 1986) ("[C]opying the claimed invention, rather than one within the public domain, is indicative of non-obviousness.").

F. The Board should deny the Petition on the merits.

For the foregoing reasons, Petitioner failed to meet its initial burden with respect to every challenged claim. Therefore, the Board should deny the Petition on the merits because Petitioner did not show "there is a reasonable likelihood that the petitioner would prevail with respect to at least 1 of the claims challenged in the petition." 35 U.S.C. § 314(a).

IV. THE PETITION DOES NOT COMPLY WITH THE IPR STATUTE OR REGULATIONS

A. The Petition does not set forth an adequate claim construction.

35 U.S.C. § 312(a)(4) provides that an IPR petition "may be considered ***only if***—"the petition provides such other information as the Director may require by

---

[7] CUDA is Petitioner's parallel computing platform and programming model that enables using one or more GPGPUs for general purpose computing.

"Schreiner1," and "¶216" and copying text as images on pages 22, 45, and 48 of the Petition. These instances add up to about 600 excess words. The Board has rejected briefs because they used similar formatting tricks to exceed page or word limits. *See, e.g., Starbucks Corp. v. Ameranth, Inc.*, CBM2015-00091, Paper 16 at 2-3 (P.T.A.B. Jan. 29, 2016) (rejecting a Patent Owner Response that used "at least 28 point" spacing rather than "double spacing").

V. CONCLUSION

For the foregoing reasons, the Board should deny the Petition.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: February 10, 2021

By: /Ted M. Cannon/
Jon W. Gurka, Reg. No. 44,139
Ted M. Cannon, Reg. No. 55,036
Cheryl T. Burgess, Reg. No. 55,030

Attorneys for Patent Owner
Advanced Cluster Systems, Inc.

CERTIFICATE OF TYPE-VOLUME LIMITATIONS UNDER 37 C.F.R. § 42.24

Pursuant to 37 C.F.R. § 42.24(d), Counsel for Patent Owner Advanced Cluster Systems, Inc. hereby certifies that this document complies with the type-volume limitation of 37 C.F.R. § 42.24(a)(1)(i). According to Microsoft Word for Office 365 word count, this document contains 11,942 words, including any statement of material facts to be admitted or denied in support, and excluding the table of contents, table of authorities, mandatory notices under § 42.8, exhibit list, certificate of service or word count, or appendix of exhibits or claim listing.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: February 10, 2021

By: /Ted M. Cannon/
Jon W. Gurka, Reg. No. 44,139
Ted M. Cannon, Reg. No. 55,036
Cheryl T. Burgess, Reg. No. 55,030

Attorneys for Patent Owner
Advanced Cluster Systems, Inc.

CERTIFICATE OF SERVICE

I hereby certify that a true and correct copy of **PATENT OWNER PRELIMINARY RESPONSE** and accompanying **EXHIBITS 2001 – 2002 and 2007 -2029** are being served on February 10, 2021, via email pursuant to 37 C.F.R. § 42.6(e), to counsel for Petitioners as addressed below:

Brent Yamashita (Lead Counsel)
Jonathan Hicks
DLA Piper LLP (US)
2000 University Ave.
East Palo Alto, CA 94303
Phone: 650-833-2348

Email: NVIDIA-ACS-IPR@us.dlapiper.com

Dated: February 10, 2021

By: /Ted M. Cannon/
Jon W. Gurka, Reg. No. 44,139
Ted M. Cannon, Reg. No. 55,036
Cheryl T. Burgess, Reg. No. 55,030

Attorneys for Patent Owner
Advanced Cluster Systems, Inc.

33919109